Original Article

External validation and update of prediction models for unfavorable outcomes in hospitalized patients with COVID-19 in Japan

Gen Yamada [a, *], Kayoko Hayakawa [a, b], Yusuke Asai [b], Nobuaki Matsunaga [b], Hiroshi Ohtsu [c], Masayuki Hojo [d], Masao Hashimoto [d], Kentaro Kobayashi [e], Ryo Sasaki [e], Tatsuya Okamoto [f], Yasuaki Yanagawa [g], Daisuke Katagiri [h], Mari Terada [b, c], Michiyo Suzuki [a], Lubna Sato [a], Yusuke Miyazato [a], Masahiro Ishikane [a], Shinichiro Morioka [a], Sho Saito [a], Norio Ohmagari [a, b]

[a] Disease Control and Prevention Center, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[b] AMR Clinical Reference Center, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[c] Center for Clinical Sciences, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[d] Department of Respiratory Medicine, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[e] Department of Emergency Medicine and Critical Care, Center Hospital of the National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[f] Department of Intensive Care Medicine, Center Hospital of the National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[g] AIDS Clinical Center, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan
[h] Department of Nephrology, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo, 162-8655, Japan

ARTICLE INFO

ABSTRACT

Introduction: Most of the currently used prognostic models for COVID-19 are based on Western cohorts, but it is unknown whether any are applicable to patients with COVID-19 in Japan.
Methods: This retrospective cohort study included 160 patients with COVID-19 who were admitted to the National Center for Global Health and Medicine between January 26, 2020 and July 25, 2020. We searched PubMed for prognostic models for COVID-19. The predicted outcome was initiation of respiratory support or death. Performance of the candidate models was evaluated according to discrimination and calibration. We recalibrated the intercept of each model with our data. We also updated each model by adding β2-microglobulin (β2MG) to the model and recalculating the intercept and the coefficient of β2MG.
Results: Mean patient age was 49.8 years, 68% were male, 88.7% were Japanese. The study outcomes occurred in 15 patients, including two deaths. Two-hundred sixty-nine papers were screened, and four candidate prognostic models were assessed. The model of Bartoletti et al. had the highest area under receiver operating characteristic curve (AUC) (0.88; 95% confidence interval 0.81—0.96). All four models overestimated the probability of occurrence of the outcome. None of the four models showed statistically significant improvement in AUCs by adding β2MG.
Conclusions: Our results suggest that the existing prediction models for COVID-19 overestimate the probability of occurrence of unfavorable outcomes in a Japanese cohort. When applying a prediction model to a different cohort, it is desirable to evaluate its performance according to the prevalent health situation in that region.

1. Introduction

Since the start of the Coronavirus disease 2019 (COVID-19) outbreak, as of November 1, 2020, there have been 46.504 million

* Corresponding author.
  E-mail address: gyamada@hosp.ncgm.go.jp (G. Yamada).

infections and more than 120,000 deaths worldwide [1], with 100,618 infections and 1773 deaths in Japan [2].

Some patients with COVID-19 become severely ill, requiring ICU management and invasive ventilation, or might die. The risk factors for severe illness have been identified, and a number of prediction models for severe disease have been developed using these risk factors [3]. Prognostic models are useful for (1) making decisions about treatment strategies, (2) allocating medical resources appropriately, (3) selecting appropriate subjects for clinical research, and (4) comparing treatment outcomes among institutions, and thus, are important from both clinical and research perspectives [4].

In a systematic review of 145 prediction models related to COVID-19, it was pointed out that many of the models are at high-risk of bias, and need to be updated before being used in local settings. For external validation, discrimination and calibration of these models should also be evaluated [3].

Most of the currently used prediction models are based on Western cohorts, but it has been suggested that the background and prognosis of Japanese patients with COVID-19, such as underlying diseases and prognosis, are different from those in Western countries [5]. Since direct application of models based on overseas data to Japanese patients might result in over- or underestimation of the probability of occurrence of the outcome, updation of the model based on local settings is necessary. In addition, although recent studies have suggested that urinary β2-microglobulin (β2MG) might be useful in predicting the severity of COVID-19 [6], to the best of our knowledge, no prediction model that includes β2MG has been investigated to date.

The purpose of this study was to examine the external validity of the existing prognostic models for COVID-19 in confirmed COVID-19 patients admitted to a tertiary care center in Japan, and to recalibrate the models. Furthermore, we aimed to update each model by adding urinary β2MG as a new predictor in the model.

2. Methods

2.1. Subjects

This retrospective cohort study included patients aged 18 years or older, admitted to the National Center for Global Health and Medicine (NCGM) between January 26, 2020, and July 25, 2020, who were diagnosed with COVID-19 by reverse transcriptase polymerase chain reaction (RT-PCR) testing. Patients who were transferred from other hospitals to our hospital for admission were excluded from the study. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Review Committee of NCGM (NCGM-G-003494-0). Information regarding opting out of our study is available on the registry website.

2.2. Measurements

Data from the COVID-19 registry Japan (COVIREGI-JP) of the NCGM were used in this study. The study data were collected and managed using REDCap (Research Electronic Data Capture), a secure, web-based data capture application hosted at the JCRAC (Joint Center for Researchers, Associates and Clinicians) data center of the NCGM [7].

Data included for the analyses were patient background factors, symptoms & signs, blood and urine test results on admission, chest radiographs on admission, and vital signs on admission. Blood samples for β2MG assessment were collected between 11 February 2020 and 25 July 2020. Since most of the subjects in this study were Japanese, obesity was defined as body mass index (BMI) $\geq$25 kg/m^2 based on the cutoff value for obesity for Asians proposed by the World Health Organization (WHO) [8]. Immunosuppression was defined by the presence of any of the following criteria: neutropenia, steroid use within the past month, chemotherapy within the past 3 months, blood transplantation, solid organ transplantation, use of immunosuppressive drugs within the past 3 months, asplenia syndrome, and primary immunodeficiency syndrome. Since direct bilirubin was not routinely measured in the data used in this study, total bilirubin was used instead of direct bilirubin when required in the calculation of the prediction model.

2.3. Outcome

The predicted outcome was initiation of respiratory support (continuous positive airway pressure/bilevel positive airway pressure, high flow nasal oxygen, invasive mechanical ventilation or extra-corporeal membrane oxygenation) or death. This corresponds to a score of 6 or higher on the WHO clinical progression scale for COVID-19 proposed by the WHO Working Group [9]. All patients were followed until discharge or death.

2.4. Search strategy for the prediction model

We searched the literature using the following procedure:
Step 1: We searched PubMed for articles published in English by November 1, 2020. The main keywords were COVID-19, prediction model, prognostic model, and logistic regression. Details of the keywords searched are described in the Supplementary data section. In addition, we hand searched several journals related to infectious diseases and emergency medicine. Step 2: We selected articles whose outcome was severe illness or death in hospitalized COVID-19 patients. Step 3: We selected articles that presented the intercept and coefficient β of the logistic regression model. If the intercept or coefficient β was not provided in the article, we requested the first author of the article to provide the missing information by e-mail. As a result, four models were selected.

2.5. Statistical analysis

Summary statistics of the patients' background factors at admission were calculated. Mean (SD), median (interquartile range), or percentage (%) values were used as appropriate.

Discrimination of each model was examined by the area under the receiver operating characteristic (ROC) curve (AUC, area under the curve). Calibration was evaluated using a calibration plot with the predicted probability on the x-axis and the observed probability
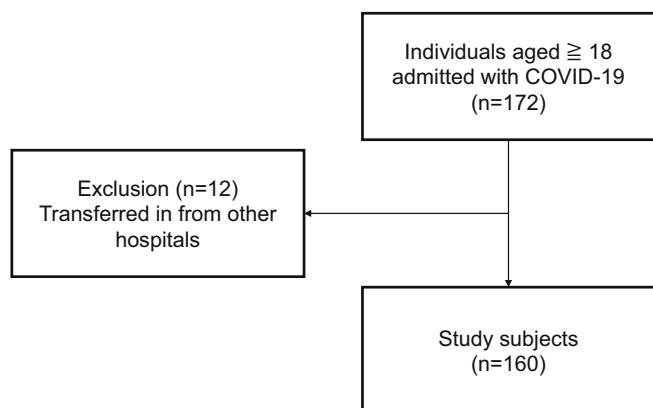

Fig. 1. Process of inclusion of study participants.

Table 1
Patient characteristics on admission.

Characteristic	Cases with available data	Total	Critical illness	
			No	Yes
			(N = 145)	(N = 15)
Days from symptom onset to hospitalization	155	7 (4–9)	6.5 (4–9)	7 (3–10)
Demographics				
Age (yr), mean (SD)	160	49.8 (18.8)	48.1 (18.1)	65.9 (17.4)
Sex (male), n (%)	160	109 (68.1)	97 (66.9)	12 (80)
BMI (kg/m^2), mean (SD)	157	23.8 (4.3)	23.6 (4.3)	25.2 (3.4)
Ethnicity	159			
Japanese, n (%)		141 (88.7)	127 (88.2)	14 (93.3)
Asian excl. Japanese, n (%)		11 (6.9)	11 (7.6)	0 (0)
White, n (%)		6 (3.8)	5 (3.5)	1 (6.7)
Others, n (%)		1 (0.6)	1 (0.7)	0 (0)
Smoking status	140			
Current, n (%)		39 (27.9)	36 (28.8)	3 (20)
Former, n (%)		34 (24.3)	28 (22.4)	6 (40)
Never, n (%)		67 (47.9)	61 (48.8)	6 (40)
Alcohol	136			
Excessive, n (%)		33 (24.3)	28 (23.0)	5 (35.7)
Sometimes, n (%)		64 (47.1)	59 (48.4)	5 (35.7)
None, n (%)		39 (28.7)	35 (28.7)	4 (28.6)
Vital signs at hospitalization				
Heart rate (/min), mean (SD)	160	85.7 (14.4)	84.7 (13.3)	96 (20.4)
Respiratory rate (/min), mean (SD)	152	18.8 (4.2)	18 (2.6)	26.4 (8.1)
Systolic blood pressure (mmHg), mean (SD)	160	123.1 (16.1)	122.8 (15.6)	126 (21.1)
Diastolic blood pressure (mmHg), mean (SD)	160	75.2 (12.6)	75.1 (12.8)	76.6 (11)
Glasgow coma scale, mean (SD)	155	15 (0.2)	15 (0.1)	14.9 (0.5)
SpO$_2$ on ambient air (%), mean (SD)	160	96.3 (2.7)	96.6 (2.4)	93 (2.9)
Laboratory data at hospitalization				
White blood cells (10^9/L), median (IQR)	156	4.96 (3.93–6.27)	5.03 (3.92–6.14)	4.71 (3.98–7.22)
Neutrophils (10^9/L), median (IQR)	146	3.39 (2.46–4.53)	3.32 (2.37–4.46)	3.71 (2.56–6.29)
Lymphocytes (10^9/L), median (IQR)	147	1.12 (0.79–1.43)	1.16 (0.86–1.51)	0.71 (0.51–0.94)
Hemoglobin (g/dL), median (IQR)	156	14.3 (12.9–15.4)	14.3 (13.0–15.3)	13.3 (11.3–15.8)
Platelets (10^9/L), median (IQR)	156	195 (160–253)	198 (163–259)	123 (94–216)
Albumin (g/dL), mean (SD)	147	3.8 (0.6)	3.9 (0.6)	3.4 (0.5)
Total bilirubin (mmol/L), median (IQR)	154	9.2 (6.8–10.3)	9.0 (6.8–10.3)	11.4 (6.8–13.7)
AST (IU/L), median (IQR)	155	31 (22–46)	30 (21.5–41.5)	44 (31–86)
ALT (IU/L), median (IQR)	156	27 (18–45)	27 (18–45)	29 (18–51)
LDH (IU/L), median (IQR)	156	233.5 (181.5–316.5)	227 (175–291)	364 (251–475)
CRP (mg/dL), median (IQR)	155	2.69 (0.43–8.57)	2.03 (0.36–6.88)	10.41 (4.81–23.22)
BUN (mg/dL), mean (SD)	153	14.5 (7.0)	13.7 (6.0)	22 (11.2)
Creatinine (mg/dL), mean (SD)	156	0.9 (0.7)	0.8 (0.7)	1.2 (0.6)
Sodium (mmol/L), median (IQR)	156	139 (136.5–142.0)	140 (137–142)	138 (133–141)
Potassium (mmol/L), median (IQR)	156	3.95 (3.6–4.2)	4 (3.6–4.2)	3.8 (3.6–4.2)
Underlying disease				
Myocardial infarction, n (%)	160	1 (0.6)	0 (0)	1 (6.7)
Congestive heart failure, n (%)	160	1 (0.6)	1 (0.7)	0 (0)
Peripheral vascular disease, n (%)	160	1 (0.6)	0 (0)	1 (6.7)
Cerebrovascular disease, n (%)	160	5 (3.1)	5 (3.4)	0 (0)
Hemiplegia, n (%)	160	3 (1.9)	3 (2.1)	0 (0)
Dementia, n (%)	160	6 (3.8)	5 (3.4)	1 (6.7)
Asthma, n (%)	160	9 (5.6)	9 (6.2)	0 (0)
Liver disease, n (%)	160	2 (1.3)	2 (1.4)	0 (0)
Diabetes, n (%)	160	22 (13.8)	18 (12.4)	4 (26.7)
Obesity (BMI≥25 kg/m^2), n (%)	157	56 (35.7)	49 (34.5)	7 (46.7)
Solid tumor, n (%)	160	4 (2.5)	4 (2.8)	0 (0)
Collagen disease, n (%)	160	3 (1.9)	3 (2.1)	0 (0)
HIV, n (%)	160	5 (3.1)	5 (3.4)	0 (0)
COPD, n (%)	160	3 (1.9)	3 (2.1)	0 (0)
Hypertension, n (%)	160	34 (21.3)	28 (19.3)	6 (40)
Hyperlipidemia, n (%)	160	25 (15.6)	21 (14.5)	4 (26.7)
Immune suppression, n (%)	160	2 (1.3)	2 (1.4)	0 (0)
Chronic kidney disease, n (%)	156	31 (19.9)	23 (16.3)	8 (53.3)
Baseline medication				
ACE inhibitor, n (%)	157	2 (1.3)	2 (1.4)	0 (0)
ARB, n (%)	157	17 (10.8)	12 (8.4)	5 (35.7)
Oral anticoagulant, n (%)	159	2 (1.3)	1 (0.7)	1 (6.7)
Antiplatelet, n (%)	159	8 (5)	7 (4.9)	1 (6.7)
Symptoms at hospitalization				
Fever≥38 °C, n (%)	160	41 (25.6)	34 (23.4)	7 (46.7)
Cough, n (%)	160	93 (58.1)	85 (58.6)	8 (53.3)
Sputum, n (%)	160	34 (21.3)	30 (20.7)	4 (26.7)
Hemoptysis, n (%)	160	2 (1.3)	2 (1.4)	0 (0)

(continued on next page)

Table 1 (*continued*)

Characteristic	Cases with available data	Total	Critical illness	
			No	Yes
			(N = 145)	(N = 15)
Sore throat, n (%)	160	39 (24.4)	36 (24.8)	3 (20)
Runny nose, n (%)	160	23 (14.4)	20 (13.8)	3 (20)
Wheeze, n (%)	160	3 (1.9)	3 (2.1)	0 (0)
SOB, n (%)	160	48 (30)	41 (28.3)	7 (46.7)
Chest pain, n (%)	160	9 (5.6)	9 (6.2)	0 (0)
Muscle pain, n (%)	160	24 (15)	23 (15.9)	1 (6.7)
Headache, n (%)	160	30 (18.8)	29 (20)	1 (6.7)
Altered mental status, n (%)	160	3 (1.9)	2 (1.4)	1 (6.7)
Fatigue, n (%)	160	73 (45.6)	65 (44.8)	8 (53.3)
Abdominal pain, n (%)	160	6 (3.8)	6 (4.1)	0 (0)
Vomiting, n (%)	160	10 (6.3)	9 (6.2)	1 (6.7)
Diarrhea, n (%)	160	27 (16.9)	25 (17.2)	2 (13.3)
Loss of taste, n (%)	160	36 (22.5)	35 (24.1)	1 (6.7)
Loss of smell, n (%)	160	27 (16.9)	26 (17.9)	1 (6.7)
Conjunctivitis, n (%)	160	1 (0.6)	1 (0.7)	0 (0)
Treatments	160			
Steroid therapy	160	29 (18.1)	15 (10.3)	14 (93.3)
Favipiravir	160	6 (3.8)	3 (2.1)	3 (20)
Ciclesonide	160	6 (3.8)	5 (3.4)	1 (6.7)
Lopinavir and ritonavir	160	9 (5.6)	7 (4.8)	2 (13.3)
Hydroxychloroquine	160	31 (19.4)	24 (16.6)	7 (46.7)
Tocilizumab	160	4 (2.5)	3 (2.1)	1 (6.7)
Remdesivir	160	7 (4.4)	4 (2.8)	3 (20)
Interferon	160	1 (0.6)	0 (0)	1 (6.7)

Abbreviations: BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; LDH, lactate dehydrogenase; IQR, interquartile range; CRP, C-reactive protein; BUN, blood urea nitrogen; HIV, human immunodeficiency virus; COPD, chronic obstructive pulmonary disease; ACE, angiotensin converting enzyme; ARB, angiotensin II receptor blocker; SOB, shortness of breath.
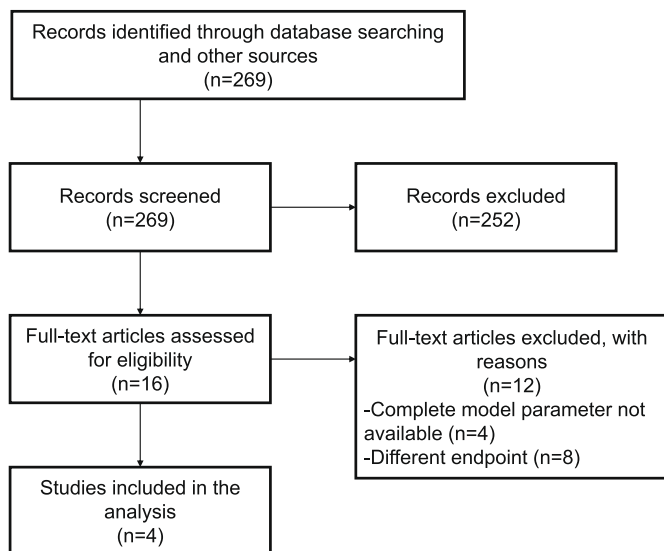

Fig. 2. Process of inclusion of previous studies.

on the y-axis. The calibration plot is a graphical representation of the actual observed probability against the model's predicted probability; if the model's prediction of the outcome probability is correct, it will overlap on the line y = x. The calibration plot was smoothed using the locally estimated scatterplot smoothing (LOESS) method. To update the models, we first recalibrated the intercept of each model with our data [10]. Next, to examine the effect of adding urinary β2MG to the model, we first selected patients with β2MG data and created a subpopulation. Using this subpopulation, we updated each model by adding β2MG to the existing prediction model and recalculating only the intercept and the coefficient of β2MG. β2MG was treated as a binary variable and

the cutoff value was set at 2457 μg/dL, as previously suggested [6]. We re-evaluated the performance of these updated models and tested whether there was a difference in AUC before and after the update.

Since this was an observational study, all available patient data were used in the analysis. All reported P values were two-sided P values, and the statistical significance level was set at $P < 0.05$. All analyses were performed with SAS software, version 9.4 (SAS Institute, Cary, NC).

3. Results

3.1. Patient background characteristics at admission

During the study period, 172 patients were included in the study, of whom 12 patients who were transferred from other hospitals were excluded and 160 patients were included in the final analysis (Fig. 1). The background characteristics of the patients at the time of admission are shown in Table 1. Mean patient age was 49.8 years, 68% were male, the majority were Japanese (88.7%), and 95.6% were Asian including Japanese. The median time from onset to hospitalization was 7 days. The study outcomes occurred in 15 patients, including 2 deaths.

3.2. Summary of the selected prediction models

We screened 269 papers by title and abstract. As a result, 16 papers were screened for full-text and four prediction models were finally selected (Fig. 2). A summary of the selected models is shown in Table 2.

3.3. Discriminatory power of the prediction models

Fig. 3 shows the ROC curves and AUCs of the four prediction models. The model of Bartoletti et al. [12] had the highest AUC (0.88; 95% confidence interval [CI] 0.81–0.96). Supplementary

Table 2
Summary of the existing prediction models.

Authors	Country	Study subjects	Predictors	Model outcome
Liang et al. [11]	China	Inpatients with COVID-19	Chest X-ray abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities, cancer history, neutrophil to lymphocyte ratio, lactate dehydrogenase, direct bilirubin,	Composite of admission to the ICU, invasive ventilation, or death
Bartoletti et al. [12]	Italy	Inpatients with COVID-19	Age ≥ 70 years, obesity, fever $\geq 38\,^\circ$C, RR ≥ 22/min, lymphocytes $\leq 0.9 \times 10^9$/L, CRP ≥ 10 mg/dL, LDH ≥ 350 IU/L, Creatinine ≥ 1 mg/dL	SpO$_2$ <93% with 100% FiO$_2$, respiratory rate >30/min, or respiratory distress
Salto-Alejandre et al. [13]	Spain	Inpatients with COVID-19	SpO$_2$ <95%, neutrophil count >7.5 $\times 10^9$/L, LDH $\geqq 300$ UI/L, CRP ≥ 100 mg/L, platelet count <130 $\times 10^9$/L	Composite of admission to the ICU or death
Ryan et al. [14]	USA	Patients who presented to the emergency department with COVID-19	Age, male sex, dyspnea, immunocompromised, chronic kidney disease, hyperlipidemia	Composite of admission to the ICU, invasive ventilation, death, or discharge to hospice

Abbreviations: COVID-19, Coronavirus disease 2019; RR, respiratory rate; CRP, C-reactive protein; LDH, lactate dehydrogenase; ICU, intensive care unit.
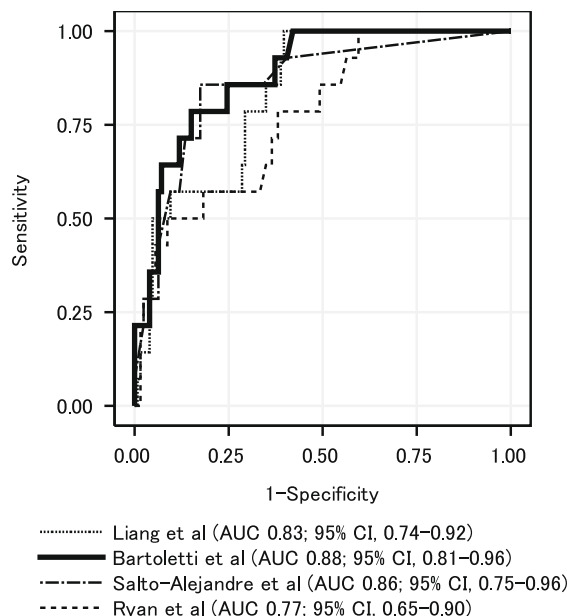

Liang et al (AUC 0.83; 95% CI, 0.74–0.92)
Bartoletti et al (AUC 0.88; 95% CI, 0.81–0.96)
Salto-Alejandre et al (AUC 0.86; 95% CI, 0.75–0.96)
Ryan et al (AUC 0.77; 95% CI, 0.65–0.90)

Fig. 3. Receiver operating characteristic curve and area under the curve of the four previous prediction models fitted to the study cohort.

Table 1 shows the sensitivity, specificity, positive predictive value, and negative predictive value at different cutoff values with application of Bartoletti et al.'s model to our cohort.

3.4. Calibration plots and recalibration of the prediction models

Fig. 4 shows the original calibration plots and post-recalibration plots of the four prediction models, and Supplementary Table 2 shows the newly estimated intercept. All four models overestimated the probability of occurrence of the outcome when the original version was applied to our cohort. Even when calibration of Bartoletti et al.'s model was assessed separately for the first wave (before June 1, 2020) and second wave (after June 1, 2020) of the pandemic in Japan [15], the model overestimated the probability of occurrence of the outcome in both periods (Supplementary Fig. 1). The post-recalibration model, in which the intercept was re-estimated using our data, showed that the Bartoletti et al. model adequately predicted the probability of outcome in our cohort.

3.5. Update of the prediction models by adding β2-microglobulin

Of the 160 patients, 133 had β2MG data and were eligible for analysis. Table 3 shows the AUCs of the original and updated models using the data of the 133 patients. The newly estimated intercept and regression coefficients for β2MG are shown in Supplementary Table 3. None of the four models showed statistically significant improvement in AUCs by adding β2MG to the model.

4. Discussion

In this study, we evaluated the performance of original and updated versions of four existing prediction models for the occurrence of the selected study outcomes of respiratory support and death in COVID-19 patients, using data from patients with confirmed COVID-19 admitted to a tertiary care hospital in Japan. The study was conducted at a single medical institution where
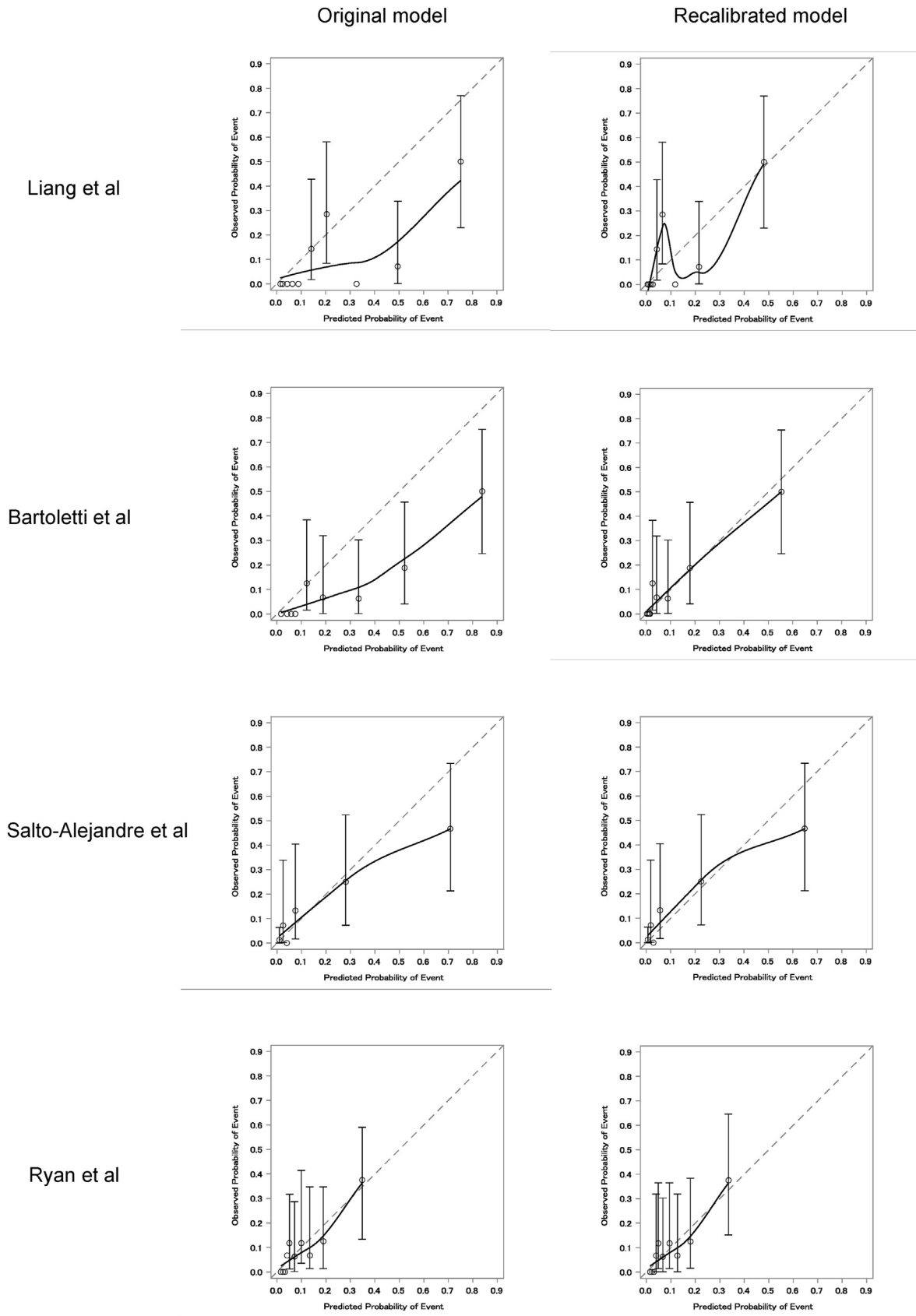
Original model   Recalibrated model

Fig. 4. Calibration plots of the four previous prediction models (original and recalibrated models).

Table 3
Area under the curve after adding β2-microglobulin to the existing prediction models.

	Original model	Updated model	P value
Liang et al. [11]	0.81 (0.70–0.91)	0.83 (0.73–0.93)	0.306
Bartoletti et al. [12]	0.87 (0.78–0.96)	0.88 (0.79–0.96)	0.447
Salto-Alejandre et al. [13]	0.84 (0.73–0.96)	0.84 (0.72–0.97)	1.000
Ryan et al. [14]	0.76 (0.64–0.88)	0.81 (0.74–0.94)	0.064

relatively uniform treatment and supportive care are provided, and therefore, accurate evaluation of the prediction model was possible. The data of this study showed that the discriminatory performance of the prediction model of Bartoletti et al. was most suitable for use in Japanese subjects, with an AUC of 0.88, and it also showed good calibration.

When calibration was evaluated, all four models calculated lower observed proportions of events than the predicted proportions of events. In other words, the prediction models overestimated the probability of occurrence of the outcomes. It has been pointed out that the rate of severe disease in Japanese patients with COVID-19 is lower than that in patients in other countries, which could be the reason why the predicted probability of the original model deviated from the observed probability in Japan [5]. The reason for the lower incidence of severe disease in the Japanese population might be fewer of the comorbidities that are reportedly associated with severe disease. For example, patients in our cohort were younger than in the other cohorts. In the studies of Bartoletti et al., Salto-Alejandre et al., and Ryan et al., the mean or median ages were 65.7, 64, and 57 years, respectively, whereas the mean age in our cohort was 49.8 years. Although data on some variables were not available in the four studies, our cohort tended to have a lower BMI and lower prevalence of hypertension, diabetes, cardiovascular disease, chronic lung disease, malignancy, and immunodeficiency. In Japan, we experienced the first wave of the pandemic (before June 1, 2020) and the second wave (after June 1, 2020), and hospitalized patients in the second wave had a shorter time between disease onset and admission than those in the first wave [15]. Moreover, since the preliminary results of the RECOVERY trial were announced on June 16, 2020, patients in the second wave in our cohort were more likely to have had a chance to receive corticosteroids, which might have contributed to the lower rate of disease severity in the second wave [16]. Another reason for the lower incidence of severe disease might be preservation of medical resources due to absence of a significant surge in the number of patients, and other factors that are not yet clear, such as genetic predisposition. The background diseases and medical situations of the population for which a model is constructed might differ greatly from those of other populations, suggesting that, as shown in this study, it is important to recalibrate the model according to the actual conditions of each region.

In this study, we assessed data from NCGM, an urban tertiary care center, that tends to include more patients with a severe condition or comorbidities as compared to other hospitals. Therefore, if the four models are evaluated in other Japanese hospitals, the models might overestimate the probability of occurrence of the outcomes even more.

When we attempted to update the four prediction models by adding β2MG, none of the four models showed statistically significant improvement. A possible explanation is that although β2MG is associated with elevated levels of inflammatory cytokines [17,18], the original prediction models already incorporated factors that reflect inflammatory cytokines, such as fever and CRP. Hence, addition of β2MG in the new model does not increase its discriminatory performance.

Many prognostic models for COVID-19 patients have been proposed [3], and in future, rather than developing new models, the external validity of existing models and updates of models should be verified. To our knowledge, this is the first such study in Japanese patients, making the methodology and findings of this study valuable. Gupta et al. externally validated 22 clinical prediction models using data from 411 inpatients in London [19]. They found that AUCs ranged from 0.56 to 0.78 in their new data and that calibration of all models was visually poor, which was consistent with the result of the present study. In the current study, we assessed four new prediction models that they did not evaluate. We also updated the models to be more suitable for predicting unfavorable outcomes in our cohort by recalibrating the models.

This study has several limitations. First, this was a single-center study, which has the advantage of accurate assessment due to the high quality of data and the relatively uniform treatment, although the number of subjects was small (n = 160), and thus, the calibration plot might not have been accurate. A study with a larger number of patients is necessary. Second, because direct bilirubin was not measured in all patients at our hospital at the time of admission, data on direct bilirubin was missing in most patients. Therefore, total bilirubin was used in evaluating the model of Liang et al., which might have introduced bias. Third, because COVID-19 is a new infection, the treatment options are still in a state of continuous flux. Patients admitted to our hospital up to July 31 were used in the analysis of this study. If treatment or prevention with significant impact on outcomes becomes available in future (e.g., widespread use of vaccines), the predicted probability of severe COVID-19 might change, in which case the clinical application of this study would require caution.

In conclusion, we present the findings of an updated prediction model that can be used to predict the severity of COVID-19 in a predominantly Japanese cohort that inherently has fewer of the known COVID-19 comorbidities. Our results suggest that existing prediction models overestimate the probability of occurrence of the outcomes of death and the need for respiratory support when used in tertiary care hospitals in Japan. Our results also suggest that when applying a prediction model to a different cohort, it is desirable to evaluate its performance according to the prevalent health situation in that region.

Funding

Authorship statement

All authors meet the ICMJE authorship criteria. All authors have seen and approved the final version of the manuscript, and contributed significantly to the work. Study concept and design: GY, KH, YA, and NM. Acquisition of data: KH, YA, NM, HO, MH, YY, DK, MT, MS, RS, YM, MI, SM and SS. Statistical analysis and interpretation of data: GY, KH, and YA. Manuscript drafting: GY, KH, and YA. Critical revision of the manuscript for important intellectual content: KH, MH, KK, RS, TO, and NO. Study supervision: NO.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jiac.2021.04.008.

References

[1] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4.

[2] Ministry of Health, Labour and Welfare. Current Status of the Novel Coronavirus Infection and the Response of the MHLW. 2020. July 8, https://www.mhlw.go.jp/stf/newpage_14573.html. accessed 2 December 2020.

[3] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

[4] Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.

[5] Matsunaga N, Hayakawa K, Terada M, Ohtsu H, Asai Y, Tsuzuki S, et al. Clinical epidemiology of hospitalized patients with COVID-19 in Japan: report of the COVID-19 REGISTRY Japan. Clin Infect Dis 2020. https://doi.org/10.1093/cid/ciaa1470.

[6] Katagiri D, Ishikane M, Asai Y, Kinoshita N, Ota M, Moriyama Y, et al. Evaluation of coronavirus disease 2019 severity using urine biomarkers. Crit Care Explor 2020;2:e0170.

[7] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inf 2009;42(2):377–81.

[8] World Health Organization Regional Office for the, Western Pacific. The Asia-Pacific perspective: redefining obesity and its treatment. Sydney: Health Communications Australia; 2000. https://apps.who.int/iris/handle/10665/206936.

[9] WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis 2020;20:e192–7.

[10] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2019.

[11] Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020;180:1081–9.

[12] Bartoletti M, Giannella M, Scudeller L, Tedeschi S, Rinaldi M, Bussini L, et al. PREDICO study group. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: a multicentre cohort study (PREDI-CO study). Clin Microbiol Infect 2020;26(11):1545–53.

[13] Salto-Alejandre S, Roca-Oporto C, Martín-Gutiérrez G, Avilés MD, Gómez-González C, Navarro-Amuedo MD, et al. A quick prediction tool for unfavourable outcome in COVID-19 inpatients: development and internal validation. J Infect 2020;S0163-4453(20). 30619-8.

[14] Ryan C, Minc A, Caceres J, Balsalobre A, Dixit A, Ng BK, et al. Predicting severe outcomes in Covid-19 related illness using only patient demographics, comorbidities and symptoms. Am J Emerg Med 2020. S0735-6757(20)30809-3.

[15] Saito S, Asai Y, Matsunaga N, Hayakawa K, Terada M, Ohtsu H, et al. First and second COVID-19 waves in Japan: a comparison of disease severity and characteristics. J Infect 2020. S0163-4453(20)30693-30699.

[16] RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in hospitalized patients with Covid-19. N Engl J Med 2021;384:693–704.

[17] Aulitzky WE, Grosse-Wilde H, Westhoff U, Tilg H, Aulitzky W, Gastl G, et al. Enhanced serum levels of soluble HLA class I molecules are induced by treatment with recombinant interferon-gamma (IFN-gamma). Clin Exp Immunol 1991;86(2):236–9.

[18] Otsubo Y, Hashimoto K, Kanbe T, Sumi M, Moriuchi H. Association of cord blood chemokines and other biomarkers with neonatal complications following intrauterine inflammation. PLoS One 2017;12:e0175082.

[19] Gupta RK, Marks M, Samuels THA, Luintel A, Rampling T, Chowdhury H, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. Eur Respir J 2020;56:2003498.